HAL open science

Cooperation in a Game of Chicken with Heterogeneous Agents: An Experimental Study

Marie-Laure Cabon-Dhersin, Nathalie Etchart-Vincent

▶ To cite this version:

Marie-Laure Cabon-Dhersin, Nathalie Etchart-Vincent. Cooperation in a Game of Chicken with Heterogeneous Agents: An Experimental Study. 2008. hal-00395939

HAL Id: hal-00395939
https://hal.science/hal-00395939

Preprint submitted on 16 Jun 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Cooperation in a Game of Chicken with Heterogeneous Agents: An experimental study

Marie-Laure Cabon-Dhersin*

Paris School of Economics
and
Ecole Normale Supérieure de Cachan.
61, Avenue du Président Wilson,
94230 Cachan, France

tel: (33)1 47 40 28 42
fax: (33)1 47 40 24 69
email: cabon@sociens.ens-cachan.fr

Nathalie Etchart-Vincent

Centre International de Recherche sur l'Environnement et le Développement (CIRED),
UMR 8568 CNRS/EHESS
Campus du Jardin Tropical,
45 bis, Avenue de la Belle Gabrielle,
94736 Nogent sur Marne Cedex, France

tel: (33)1 43 94 73 78
fax: (33)1 43 94 73 70
email: etchart@centre-cired.fr

*Corresponding author: M-L Cabon-Dhersin, Ecole Normale Supérieure de Cachan, 61 avenue du Président Wilson, 94230 Cachan, France. tel: (33)1 47 40 28 42, email: cabon@sociens.ens-cachan.fr

Cooperation in a Game of Chicken with Heterogeneous Agents: An experimental study

Abstract:

Using a two-stage Game of Chicken, Cabon-Dhersin and Ramani (2007) examine the impact of population heterogeneity on cooperative behaviour. For that purpose, they introduce two different types of agents, namely those who always cooperate and those who strategically cooperate. The aim of the present experimental study is to investigate the descriptive accuracy of the one-stage version of this model. Even though the level of cooperation appears to be much higher than predicted, our data bring strong support to the main predictions that the structure of the population as well as the structure of the payoffs affect cooperative behaviour.

Keywords: Social dilemma; Game of Chicken; Experiments; Cooperation; Heterogeneity.

Cooperation in a Game of Chicken with Heterogeneous Agents: An experimental study

Introduction

A huge body of literature has been devoted to the experimental investigation of subjects' proneness to cooperation, in contexts in which game theory predicts that they should either barely cooperate or not cooperate at all. The most investigated two- and multiple-person games (prisoner's dilemma game, trust game, public goods game, game of chicken) describe so-called 'social dilemma' situations, in which the strategic choice of each agent (who tries to maximize her own earnings) results in a non-cooperative equilibrium that makes everyone worse off (Hardin 1971). However, experimental evidence shows that the level of cooperation, even in one-shot games, is much higher than predicted by theory (Berg et al. 1995; Güth et al. 1997).

A question remains unanswered at this stage: why do people choose cooperation while they may be betrayed by their partner and do not have any monetary incentive to cooperate?

Several kinds of theoretical explanations have been proposed and empirically investigated. For instance, biological as well as ecological, anthropological and social evolution seems to be favourable to cooperation (Ridley 1997). In this respect, social identity-related factors (Dawes et al. 1988; Kollock 1998b; Yamagishi and Kiyonari 2000; Wit and Wilke 1992), social norms (Thogersen 2008), as well as socio-demographic and cultural factors (Boone and van Witteloostuijn 1999; Gächter et al. 2004) may play a role in the prevalence of cooperative behaviour. In the same vein, personality[1] can be shown to strongly affect behaviour in interactive settings, with aggressive people being less cooperative than sociable ones (Boone et al. 1999). Moreover, cheating is likely to induce psychological costs that many people prefer to avoid. More generally, some agents appear to be motivated by considerations of “fairness” and “reciprocity” that may make them opt for cooperation (Bolton and Ockenfels 2000; Charness and Rabin 2002; Falk and Fischbacher 2006; Fehr and Schmidt 1999; Levine 1998; Rabin 1993). Finally, cognitive limitations may lead some subjects to use in one-shot games the rules of thumb they have developed in repeated games (Goeree and Holt 2001), which may increase their proneness to cooperation in experimental settings.

The above-mentioned arguments may contribute to explain why people cooperate much more than they are expected to. They also remind us that strong heterogeneity can be expected to prevail among people in interactive settings. To be more specific, some recent and stimulating empirical results suggest that, even though a non negligible minority or even a majority of subjects maximize their own monetary gains, a significant proportion of them exhibits other-regarding preferences. For instance, using data from a wide rang of experiments, Fehr and Gächter (2000) (cited in Camerer 2003) estimate that 40 to 66% of the subjects show a preference for reciprocity, while only 20 to 30% of them care about their sole monetary gains (the remaining subjects do not seem to have very clearly defined preferences). A recent experimental study involving a Game of Chicken (Neugebauer et al. 2008) suggests that, even though most of the subjects (83%) tend to behave in a self-interested way and to maximize their monetary gains, some of them appear to favour reciprocity and equity.

In the recent years, heterogeneity among agents has also been incorporated into several theoretical models (e.g. Buskens 2003; Cabon-Dhersin and Ramani 2007; Frank 1987; Yaari 2001). For instance, Cabon-Dhersin and Ramani (2007) (CDR in the following) examine (among other things) how proneness to trust and cooperation of a given decision maker is to be affected by the heterogeneity of the population, i.e. by the probability that she be faced with a partner of a different behavioural type. For this purpose, the authors use a two-stage one-shot Game of Chicken framework (they also examine the case of the Prisoner's Dilemma Game, which will not be studied here). The Game of Chicken is a social dilemma that has been much less investigated than the Prisoner's Dilemma Game (PDG). It describes interactive settings in which each partner individually can behave in a way that will benefit both partners. Like in the PDG, each partner will benefit more from bilateral cooperation than from bilateral defection. However, like in the PDG, such a cooperative behaviour is costly, so it will make no sense for a "self-interested" individual to consent an effort toward cooperation if she expects her partner to cooperate. In that case, her best strategy will be defection. But, contrary to what happens in the PDG, if she expects her partner to defect, she will have interest to cooperate, and she will be declared the "chicken" of the game. Therefore, in the Game of Chicken, two pure equilibria exist (corresponding to unilateral cooperation and unilateral defection respectively), which means that, contrary to the PDG, it does not imply any dominating strategy (for some further details about this game, see for instance Kollock 1998a). The Game of Chicken is meant to capture strategic interactions between individuals, firms, institutions, social groups, political parties and countries. For instance, it has been often used to describe military or political conflict (Snyder 1971, Stone 2001), as well as negotiations over environmental conventions (Carraro and Siniscalco 1993, Ward 1993).

CDR (2007) is based on a two-person two-stage Game of Chicken with two kinds of players, called the 'opportunists' and the 'non-opportunists' respectively. In the first stage, the choice options are to initiate cooperation or not. In the second stage (provided cooperation has been initiated by both players), the options are to abide by one's commitment to cooperate or not. In that framework, the 'opportunists' seek to maximize their monetary gains; they will not choose to abide by their initial commitment to cooperate unless it is in their interest to do so. By contrast, the 'non opportunists' follow a single rule, which is to always respect their initial commitment to cooperate. Their actions are driven by some other motives than opportunism (e.g. self respect, altruism, ethical considerations). Furthermore, information incompleteness is assumed: each agent is aware of her own type, but ignorant of the type of her partner in the game. However, the distribution of the types in the population is common knowledge. Because it introduces uncertainty as regards the type of one's partner in the game, heterogeneity is likely to complicate and strongly affect the way people make their decisions.

For the sake of simplicity, and with no loss of generality, only the one-stage version of the model will be experimentally investigated here, meaning that the subjects will only have the choice between cooperating or not cooperating. In this simplified one-stage framework, the opportunists will be called 'strategic cooperators' (SCs), while the non opportunists will be called 'unconditional cooperators' (UCs).

Under the previous set of assumptions, and at the symmetrical mixed-strategy equilibrium, the main and rather counterintuitive prediction of the model is that, the higher the ratio of SCs in the population (or, in other words, the higher the probability to meet and play against a SC), the higher their probability to cooperate. Besides, a threshold effect is at play, since a minimum percentage of SCs in the population is required to induce a cooperative behaviour on their part. Another remarkable result is that, beyond the same threshold (i.e. beyond the same given proportion of SCs in the population), the degree of cooperation in the whole population does no longer depend on the proportion of SCs, but is only affected by the structure of the payoffs. Finally, the model provides some predictions about how the degree of cooperation should be affected by the structure of the payoffs.

The main purpose of the present experimental study is to investigate the descriptive accuracy of CDR's model. Using a neutral (context-free) display, we wish to examine to what extent and how both the structure of the population and the structure of the payoffs are likely to affect the level of cooperation among the subjects. The main results are twofold:

(i) *From a qualitative point of view*, subjects appear to behave in accordance with the model's predictions. First, the intensity of cooperation among the SCs as well as among the whole population appears to depend on the structure of the population in the theoretically expected way. Second, predictions as regards the sensitivity of behaviour to the structure of the payoffs are also qualitatively validated: the level of cooperation appears to be a decreasing function of the unilateral defection gain, as well as an increasing function of the unilateral cooperation gain.

(ii) *From a quantitative point of view*, the empirical proportion of cooperative choices turns out to be systematically 20 to 50 points superior to the theoretically expected one.

The theoretical framework and the predictions that our experimental study was designed to test are presented in the following section. The experimental design and procedure are described next. The last two sections are devoted to the results and discussion respectively

The theoretical framework and predictions

CDR (2007) is based on a two-person two-stage Game of Chicken with incomplete information. For the sake of simplicity, and with no loss of generality, only the one-stage version of the model will be presented here[2]. As said in the introduction, the population of the players is assumed to consist of two kinds of agents, called 'strategic cooperators' (denoted SCs in the following) and 'unconditional cooperators' (denoted UCs) respectively. The SCs are expected-value maximizers and choose not to cooperate unless it pays to do so. Consequently, the SCs can either defect, choosing action *d*, or cooperate, opting for action *c*.

By contrast, the UCs follow the simple rule always to cooperate, whether it pays to do so or not. Their only option is thus to choose cooperation and play *c*.

In the game, each player knows her own type and has some probabilistic information about the type of her partner. Let $p \in (0, 1)$ be the probability that the player's partner is a SC. This probability is exogenously given and is common knowledge to all players. Notice that p can also be interpreted as the proportion of SCs in the population. Besides, each player expresses a belief as regards the behaviour of her fellow player. The game is symmetrical, in the sense that the structure of the payoffs and the beliefs of players of the same type are assumed to be identical. For each player, her endogenous belief that her fellow player will not cooperate *if she is a SC* is given by a probability α and it lies in the interval (0, 1). So, her belief that her partner (*whatever her type*) will cooperate is given by the probability $(1-p\alpha)$[3].

The payoff structure of the game is as follows. Each player wins X if both players cooperate and Y if they both deviate. If a SC defects while her partner cooperates, she gets $H > X$ while her partner gets $L > Y$ with $L < X$. This payoff structure clearly captures the essence of the Game of Chicken. The matrix of the game is given in Figure 1.

Insert ***Figure 1.*** *about here.*

The Game of Chicken allows multiple equilibria. Since defection is not a dominant strategy, it is possible that, at least in some equilibria, cooperation prevails even among the SCs[4]. Here, we will focus on those situations in which cooperation can be observed, *i.e.* in which the SCs will choose to cooperate with a positive probability $(1-\alpha) > 0$. Besides, since the structure of the payoffs and the beliefs of players of the same type are assumed to be identical, only symmetrical equilibria will be considered.

For a SC, a necessary condition for cooperation is that her expected returns from cooperation be greater than those from defection. This condition is given by:

$$p\left[(1-\alpha)X+\alpha L\right]+(1-p)X > p\left[(1-\alpha)H+\alpha Y\right]+(1-p)H$$

In other words, we have:

$$(1-p\alpha)<T, \text{ with } T=\frac{L-Y}{(H-X)+(L-Y)}$$

Some remarks should be made at this stage. First, in a Game of Chicken, cooperation is possible but never certain. In fact, it can be shown that the exact probability with which a SC will defect is given by $\alpha=\frac{1-T}{p}$. Second, a necessary condition for cooperation to emerge is that the probability of non-defection (i.e. the probability that the partner chooses cooperation), given by $(1-p\alpha)$, be not too high. When $(1-p\alpha)$ is equal or superior to a given value *T*, i.e. when $(1-p)\geq T$ (with $\alpha<1$), it is always optimal for a SC to deviate (The details of the proofs are given in CDR, 2007).

Result*: In the one-shot one-stage Game of Chicken under consideration, the symmetrical Nash equilibrium is such that:*

(i) *Whenever* $p>(1-T)$, *SCs will cooperate with a probability* (1 – α) such that:

$$(1-\alpha)=1-\frac{1-T}{p}=1-\frac{1}{p}\frac{H-X}{(H-X)+(L-Y)}.$$

(ii) Whenever $p\leq(1-T)$, *SCs will defect with a probability* $\alpha=1$.

Remembering that the a priori probability *p* that the partner be a SC may also be interpreted as the proportion of SCs in the population, the following four testable assumptions A1 to A4 can be drawn from the previous result.

A1: A minimum proportion $(1-T)$ of SCs in the population is necessary to induce a cooperative behaviour on their part.

When the population contains only a few SCs, these will always defect because they know that their partner is very likely to cooperate (see Figure 2, left part).

A2: When $p > (1-T)$, the higher the proportion p of SCs in the population, the stronger their incentive to cooperate.

When $p > (1-T)$, $\alpha = \frac{(1-T)}{p}$ at equilibrium. So, the probability α assigned by a SC to defection decreases when *p* increases (see Figure 2, right part). Note that, for *p* = 1, $\alpha = 1-T$. This corresponds to the standard mixed-strategy equilibrium when there is no heterogeneity in the population, i.e. when the population contains only expected-value maximizers (SCs).

*Insert **Figure 2** about here.*

A3: For any given configuration of payoffs and for any population with $p > (1-T)$, the proportion of cooperative agents in the whole population *is constant and equal to T.*

As long as the proportion *p* of SCs remains lower than $(1-T)$, the expected return from defection remains higher than that from cooperation, and any SC must defect whatever the payoff configuration ($\alpha = 1$). Therefore, the proportion of agents who cooperate *in the whole population* is equal to the proportion $(1-p)$ of UCs in the population (see Figure 3, left part).

When *p* becomes higher than $(1-T)$, the probability of defection (given by the probability α at equilibrium), decreases as *p* increases. Quite remarkably, the exogenous increase in the proportion *p* of SCs is exactly counterbalanced by the decrease in their endogenous propensity to defect α, so that the proportion $p\alpha$ of agents who defect *in the whole population* becomes constant and equal to $1-T$. Alternately, the proportion $(1-p\alpha)$ of agents who cooperate in the *whole population* is constant and equal to *T* (see Figure 3, right part).

*Insert **Figure 3** about here.*

A4a: The probability of defection α is an increasing function of the return from unilateral defection (H).

A4b: The probability of defection α is *a decreasing function of the return from unilateral cooperation (L).*

$T = \dfrac{L-Y}{(H-X)+(L-Y)}$ is a decreasing function of *H* and an increasing function of *L*. Since $\alpha = \dfrac{(1-T)}{p}$, α appears to be a decreasing function of *T*, thus an increasing function of *H* and a decreasing function of *L*.

Now, let us present the design of our experimental study, which aimed at testing the previous four assumptions A1 to A4.

The experimental design

85 subjects (among which 45 female) participated in the final experiment. All the subjects were between 22 and 28 years old. Most of them were undergraduate students in Economics; the remaining ones were undergraduates in Mathematics. All of them were aware of game theory and decision theory, but with no specific skills in these topics.

The experiment was run during January and February 2007, and consisted in three successive and at least 15-days spaced independent sessions. In each session, the subjects' task consisted in the filling out of a paper-and-pencil questionnaire. Only the results from the first session/questionnaire will be reported here.

The questionnaire consists of three separate parts, including several interactive choice situations each. In each choice situation, the subject is given a specific Game of Chicken payoff structure (with X, Y, H and L as in Figure 1 *supra*) as well as some probabilistic information about the type of her partner (being either a SC with a probability p or a UC with a probability $(1-p)$) and asked to decide whether to cooperate or not with her. X (bilateral cooperation gain) and Y (bilateral defection gain) are given the same values X = 100 € and Y = 50 € throughout the questionnaire.

The *first part of the questionnaire* aims at investigating whether the subjects' behaviour changes as the proportion of UCs in the population increases from 0% to 100% (or, alternately, as the probability to meet a UC grows from 0 to 1), with the unilateral defection gain H and the unilateral cooperation gain L being kept constant and given the values H = 120 € and L = 70 € respectively (Assumptions **A1**, **A2**, **A3**). The *second part of the questionnaire* is designed to investigate whether the subjects' behaviour changes as the value of the unilateral defection gain H increases from 100 € to 190 €, with p = 0.75 and L being kept constant and given the value L = 70 € (Assumption **A4a**) Finally, the *third part of the questionnaire* aims at investigating whether the subjects' behaviour changes as the value of the unilateral cooperation gain L increases from 50 € to 95 €, with p = 0.75 and H being kept constant and given the value H = 160 € (Assumption **A4b)**.

To facilitate the subject's task, each part of the folder includes a series of 3 questions which is built following the same pattern. A typical choice situation (drawn from the first part of the questionnaire) is reproduced in Appendix A. In the 1st and 2nd questions, the subject is provided with the extreme values under investigation (p = 0 and p = 1; H = 100 € and H = 190 €; L = 50 € and L = 95 € respectively). In the 3rd question, a table recapitulates the whole range of 10 or 11 values[6]; for each value in the table, the subject has to choose between cooperation and non cooperation by tipping the appropriate box. Note that the cooperative/non cooperative options are actually neutrally labelled 'red'/'blue', while the UCs (resp. SCs) are labeled 'type (a) agents' (resp. 'type (b) agents') to keep the wording as neutral as possible and avoid any framing effects, due to social and/or moral considerations for instance[5] .

The folder took about 15 minutes to be completed. The subjects were given written instructions, which were complemented with some oral instructions. In particular, the subjects were reminded that their partner would change from one question to another and that they had to consider each successive choice in isolation, especially when filling in the tables[7]. Besides, they were reminded that there were neither wrong nor right answers, and that they could switch from 'red' to 'blue' (or the opposite) as often as they wished to, or even choose to always play 'red' or always play 'blue'.

The participants were paid 5 euros for participating in the whole study (they were paid once they had completed the third questionnaire). Moreover, to help them consider their decisions as real despite the apparently fictitious nature of the choice situations, a high-incentive performance-based payment procedure was also introduced in each session. At the beginning of the session, the participants were informed that i) two of them would be selected at random at the end of the session[9], ii) each of them would be asked to draw (at random) one of the choice situations included in the questionnaire and to play it out for real *against a real partner* (namely another participant), iii) her gains would depend on her previous choice as well as on her partner's previous choice. More details as regards the instructions as well as the payment procedure are available upon request.

Note that Assumptions **A1** to **A4** could not be tested unless the subjects' type be identified. In our experiment, each subject was classified through her behaviour in a single choice situation, namely the one encapsulated in Question 1.2. ("Which colour do you choose when you have 100% chances to meet a partner of type (a), who always chooses 'red'?"). Indeed, this choice situation is the only one in which any self-interested subject is required to defect (play 'blue'). So, those subjects who chose to cooperate were labelled as non SCs, and the defective ones as SCs. Naturally, this rather weak criteria does not ensure that those latter subjects were genuine SCs. But the main point is that their behaviour throughout the questionnaire was fully compatible with that of a SC. Besides, and quite reassuringly, the rate of subjects labelled as SCs using our criterion appears to be very close to that found in Neugebauer et al. (2008)'s study involving a Game of Chicken (86,7% versus 83%).

Now, what about those subjects who chose to cooperate in Choice Situation 1.2.? As said before, those subjects were not classified as UCs, but as *non* SCs. To understand why, the argument is twofold. First, an important point to make is that we did not need to collect any empirical information about the UCs to be able to test Assumptions A1-A4. Indeed, the behaviour of a UC is unambiguous since she always cooperates. So, given the proportion $(1 - p)$ of UCs in the population (provided in each scenario of the questionnaire), the proportion of cooperative choices among the UCs could be obtained *without considering the empirical population of UCs* (it is simply equal to $(1 - p)$). Second, we are well aware that the a priori SC and UC two categories are unlikely to exhaust the components of *real* world population. By the way, most of the subjects who chose to cooperate in Choice Situation 1.2. were not genuine UCs, since they chose *not* to cooperate in some other choice situations. So it would have been fallacious to label them as UCs.

This basic ambiguity, as well as the fact that we were not really interested in the UCs, led us to exclude from the set of data that was used to investigate the descriptive accuracy of the model all the subjects whose behaviour was not fully compatible with the SC profile[10].

After discarding those subjects who appeared to produce erratic answers, we were finally left with a 72 subject sample.

Results

We successively examine the impact on cooperation of population heterogeneity (Assumptions A1 to A3) and payoff structure (Assumptions A4a and A4b).

Cooperation depending on the structure of the population

We successively discuss the impact of population heterogeneity on cooperation among the SCs (**A1** and **A2**) and among the whole population (**A3**).

First, note that, thanks to the law of large numbers, the probability α that a SC deviates (resp. the probability $(1-\alpha)$ that a SC cooperates) can be proxied as the *proportion* of non cooperative choices (resp. cooperative choices) among the *empirical* SC population.

Now, Graph 1 reports both the theoretical and empirical proportions of defective SCs depending on the proportion of SCs in the population. A proportion test shows that, for any strictly positive *p*, the difference between the empirical and theoretical levels of cooperation is always very significant and positive (every p-value is equal or inferior to 0.001). Moreover, the degree of over-cooperation increases until $p = 0.75$. To better describe the results, we now distinguish the low and high parts of the curves.

For any $p \leq (1-T) = 0.5$, **A1** predicts that the SCs should not cooperate at all. For $p = 0$, all the SCs actually defect (as a result of the criterion used to identify the SCs in our population). But when *p* grows from 0 to 0.5, the gap between the empirical and theoretical levels of cooperation grows, reaching almost 40% for $p = 0.5$.

When $p > (1-T) = 0.5$, the level of cooperation grows as *p* increases (i.e. as the probability of playing against a SC increases), as predicted by **A2**. And the theoretical and empirical curves appear to be remarkably parallel, which suggests that the model works qualitatively well when the SCs are in a majority in the population ($p > 0.5$). Nevertheless, the initial gap between the theoretical and empirical levels of cooperation remains, and is kept rather constant for $0.5 \leq p \leq 1$. It is noteworthy that, in the standard case without heterogeneity (i.e. for $p = 1$), the level of cooperation is 20% higher than expected: while no more than 50% cooperative choices should be observed among the SCs, this rate actually reaches 70%.

Insert ***Graph 1*** *about here.*

Now, as regards **A3**, note that the proportion of cooperative choices in the whole population, given by $p(1-\alpha)+(1-p)$, was computed using the *empirical* proportion $(1-\alpha)$ of cooperative choices among the SCs (identified as such using Question 1.2.), the *theoretical* proportion *p* of SCs in the population given in each scenario, and the *theoretical* complementary proportion $(1-p)$ of UCs.

A3 predicts that the proportion of cooperative choices *in the whole population* depending on the proportion of SCs in the population should decrease from 100% to 50% as *p* grows from 0 to 0.5 (see Graph 2). Then, this proportion should remain constant, the increase in the proportion of *cooperative* SCs being exactly counterbalanced by the decrease in the proportion of UCs (who always cooperate). Though not completely supported by our data from a quantitative point of view, this strong theoretical property appears to be qualitatively rather satisfied.

As regards the left part of Graph 2 (from p = 0 to p = 0.25), the theoretical and empirical curves match very well. Looking at Graph 1 may help understand why. Actually, our SCs do cooperate more than expected (remember they should not cooperate at all), but since they are in a small minority in the population, their behaviour has not much impact on global behaviour[11].

Now, from p = 0.33 to p = 0.5, cooperation decreases slowlier than it should, so that the global level of cooperation remains 20% too high. Once again, the explanation can be found in Graph 1. Indeed, the degree of over-cooperation among the SCs grows with p. Since their weight in the population increases, the degree of over-cooperation among the whole population also increases.

From p = 0.5 to p = 1, the level of cooperation remains too high (it varies between 70 and 80% instead of remaining constant at 50%). Moreover, the empirical curve exhibits a somewhat inverse-U shape instead of a flat shape, which is due to the fact that the degree of over-cooperation among the SCs tends to decrease as p grows. Since their weight in the population increases, the aggregate level of over-cooperation is also affected downward.

Anyway, a striking and noteworthy result is that the level of cooperation observed for $p = 1$ is equal to its value for p = 0.5, which fits the theoretical prediction and suggests that the difference between predicted and observed behaviour is essentially quantitative: on the one hand, the level of cooperation does not dramatically decrease as it should, and it always exceeds the predicted level by about 20 points; but, on the other hand, the empirical curve appears to be quite remarkably parallel to the theoretical one.

*Insert **Graph 2** about here.*

Cooperation depending on the structure of the payoffs

The level of cooperation has been shown experimentally to depend on the structure of the payoffs (see for instance Sherman, 1969 or Güth et al., 1997). CDR (2007) predicts how the structure of the payoffs should affect behaviour (Assumptions **A4a** and **A4b**). Our experimental study precisely allows us to investigate A4a and A4b and to identify the role played by *H*, the unilateral defection gain, and by *L*, the unilateral cooperation gain.

First, the level of cooperation is expected to decrease as *H* grows (**H4a**), with a maximum level of cooperation of 100% for *H* = 100 €, and a flat right part and no cooperation at all when *H* becomes superior or equal to 160 €.

The level of cooperation actually appears to decrease as *H* grows. Once again, the empirical level of cooperation appears to be quantitatively higher than expected by 30 to 50 points, and this difference in level is highly significant for all *H* (z test; every p-value is equal or inferior to 0.001). But, on the other hand, the empirical curve appears to be qualitatively similar to the theoretical one (see Graph 3).

Quite remarkably, the percentages obtained for the last 4 values (*H* = 160 €, 170 €, 180 €, and 190 €) appear not to be significantly different (proportion test, p-value: 0.754). The level of cooperation can thus be considered as constant between *H* = 160 € and *H* = 190 €, which fits the theoretical prediction. However, the cooperation curve is still strongly translated upward: while for any $H \geq 160$ €, theory predicts that the SCs should no longer cooperate, the level of cooperation actually remains strongly positive (about 25/30%), even for the very deterrent gain of 190 €.

*Insert **Graph 3** about here.*

Now, the level of cooperation is expected to increase as L grows (**H4b**), with a flat left part and no cooperation at all as long as L is inferior or equal to 70 €, and a maximum level of cooperation of about 25% for L = 95 €.

The level of cooperation actually appears to increase as L grows. However, as previously, the observed level of cooperation appears to be much higher than predicted (with all p-values equal or inferior to 0.001). Three striking features are worth noticing. First, the left part of the empirical curve ($L \leq 60$ €) is flat, as predicted by the model. There is naturally a quantitative difference, since some cooperation exists while there should be no cooperation at all. But it seems that, as expected, proneness to cooperation is not sensitive to variations of L when L is kept small. As a second important feature, the level of cooperation dramatically increases when L grows from 60 to 80 €, as if the subjects became highly sensitive to variations of L when L is given intermediate values, with a gap reaching 50% for L = 80 €. Thirdly, and contrary to theoretical predictions, the right part of the empirical curve appears to be flat, as if the subjects became insensitive to variations of L when L is given high values (from 80 to 95 €). A proportion test run on the rate of non cooperative choices for these 4 values gives a p-value of 0.964.

*Insert **Graph 4** about here.*

Discussion

Methodological discussion: the experimental design

Part of the gap between the predicted and observed levels of cooperation might be explained by the fact that it was actually impossible to fully replicate experimentally the conditions under which the theoretical predictions were obtained. To be specific, two main assumptions had to be made as regards the behaviour of *real* subjects when designing the experiment to allow for the testing of the model. Since it is actually impossible to know whether these assumptions were experimentally satisfied or not, they might be viewed as undesirable 'auxiliary hypotheses', raising a Duhem-Quine argument and preventing us from drawing clear-cut conclusions from our data (see for instance Starmer, 1999). In the following paragraph, we will present these two assumptions, as well as some empirical arguments allowing us to remain quite confident about our data.

The first assumption that was implicitly made in our experimental design regards the nature of the equilibrium played by the subjects – namely a mixed-strategy equilibrium. Indeed, the four predictions A1 to A4 we aimed at investigating were obtained at mixed-strategy equilibrium. But the model itself offers multiple equilibria, and we can actually not be sure that our subjects coordinated toward the mixed-strategy equilibrium. Furthermore, CDR (2007)'s theoretical results were obtained under the assumption of symmetry, under which the agents are assumed to consider that any partner of the same type should have the same beliefs as theirs. Obviously, this may not be true in the real world. In our experiment, it may have been the case that some of our subjects considered that their fellow player would have different beliefs and would play a different equilibrium.

Though theoretically valid, we think that this argument can be dismissed using several empirical counter-arguments. First, the qualitative similarity between the theoretical and empirical curves (see Graphs 1 to 4, Section 4 supra) suggests that the experiment succeeded quite well in capturing the qualitative features of the model. Second, even though the subjects did actually not elaborate mixed strategies, the fact that aggregate behaviour roughly coincides with mixed-strategy equilibrium behaviour gives some weight to the *as if* hypothesis, which has received some support in economics (Friedman and Savage, 1948). By the way, mixed strategies can also be interpreted as the aggregate result of the mix of different pure strategies among the subjects.

The second assumption we wish to discuss here concerns the categorization of the population. The questionnaire implicitly followed CDR (2007) in assuming that the population of subjects be divided into two classes of agents, namely the SCs and the UCs. As already mentioned in Section 3, we are well aware that these a priori categories are unlikely to exhaust the components of *real* world population, and that some other behavioural types could be identified (for instance a category of ‘unconditional defectors’, who always choose to defect; see Neugebauer et al., 2008 for an example of a four-type categorization). Obviously, if some subjects in our experimental study turned out to be of another type, this would prevent us from testing the two-type model under consideration. However, several counter-arguments can be put forward, that minimize the empirical weight of the previous argument. First, in the pilot experiment as well as in the final experiment, none of the subjects showed either disagreement or surprise toward the two-type categorization when filling out the questionnaire. Furthermore, remember that only those subjects whose behaviour was consistent with the behaviour of a SC were retained for data analysis. Maybe they were actually something else, but for our purpose, only their *behaviour* matters. Besides, the subjects whose behaviour was inconsistent with expected-value maximizing were simply considered as non-SCs. Implicitly, this allowed us to take into account the possibility that some subjects be neither SCs nor UCs. And to keep the data as unambiguous as possible, these subjects were not included in the final data set.

Discussion of the results

The influence of social determinants

Even though our experimental design may have introduced some noise in the data, it is unclear how the simplifications we made could be responsible for the high level of cooperation observed among the subjects.

However, such factors as personality as well as social or equity considerations (reviewed in the introduction) may have obviously contributed to our results. For instance, a huge amount of recent economic literature suggests that, contrary to what basic theory suggests, people do care about the way gains are shared out between them and others (McClintock and Liebrand 1988) and tend to exhibit social preferences such as preference for reciprocity or inequity aversion (see Fehr and Fischbacher 2002 for a survey). Preference for reciprocity entails that actions that are perceived to be kind (resp. hostile) will be reciprocated in a kind (resp. hostile) manner (Rabin 1993; Levine 1998; Charness and Rabin 2002). *Inequity aversion* can be viewed as a kind of altruism: inequity-averse people will try to promote their partner’s earnings – provided these earnings remain acceptable as compared to theirs (Fehr and Schmidt 1999; Bolton and Ockenfels 2000). In this framework, even selfish people may be induced to make ‘non-selfish’ choices. This may contribute to explain why our subjects cooperated much more than predicted in CDR (2007).

Demographic as well as socio-cultural determinants may have also played some role in our results. For instance, social identity has been shown to promote cooperative behaviour among in-group members (Dawes et al. 1988; Kollock 1998b; see also Yamagishi and Kiyionari, 2000 and Simpson, 2006 for some recent work on the links between social identity and cooperation), due to ‘the shared and mutual perception by in-group members of their interests as interchangeable’ (Turner et al. 1987). Since our sample was highly homogeneous as regards age and human capital (all of them were graduate students), as well as social group identity (most of them were students in a French 'Grandes Ecole'), it may be the case that all these factors combined to induce highly cooperative behaviour.

To complete this discussion, we would like to raise another possible explanation for the high level of cooperation observed among our subjects. This explanation has to do with the subjects' attitude toward risk.

The role of risk attitude.

As most game theoretical models, CDR (2007) assumes that the SCs maximize the expected value of the game. What we wish to do now is to relax this core assumption by introducing more sophisticated preferences under risk, allowing the subjects to subjectively deal with outcome and probability in a non linear fashion instead of simply maximizing expected value.

Intuitively, the basic uncertainty that is associated with the partner's behaviour (Sabater-Grande and Georgantzis, 2002) makes it natural to consider interactive situations as a kind of risky situation – involving risky payoffs with a probability distribution over the payoffs. In this perspective, the decision to cooperate (or trust one's partner) may be viewed by the subject as equivalent to taking a risky bet (Bohnet and Zeckhauser 2004, Eckel and Wilson 2004). If so, her risk orientation is likely to influence her 'cooperative attitude'. In the Game of Chicken (and contrary to what happens in the PDG), the cooperative strategy (action *c*, see Figure 1 supra) appears to be less risky than the non cooperative one (action *d*), since it allows to avoid the worst payoff in the worst outcome (bilateral defection). So, more risk aversion can be expected to result in more cooperation. In CDR (2007), the risky nature of the game is all the more obvious since it introduces a second kind of uncertainty through the probabilistic type of the partner. The decision maker has to think about both the type (SC or

UC) and the behaviour (cooperation or defection) of her partner. To our knowledge, only a few experimental studies have been trying to explicitly link together attitude toward risk and attitude toward cooperation (Bohnet and Zeckhauser 2004; Brennan et al. 2007; Dolbear and Lave 1966; Eckel and Wilson 2004; Sabater-Grande et Georgantzis 2002). Unfortunately, no conclusive result can been drawn from these quite contradictory studies. Part of the apparent contradiction may come from the fact that attitude toward risk strongly depends on the way it is elicited. For instance, the subjects' degree of risk aversion turns out to be higher when elicited in an interactive context (in a social dilemma for instance) than in an exogenously risky context (through the comparison of lotteries for instance) (Bohnet and Zeckhauser 2004). This may obviously confuse the investigation of the connections between attitude toward risk and proneness to cooperation.

Anyway, a huge body of theoretical as well as empirical literature in the field of individual decision making under risk strongly confirms that most people are not risk neutral and do not treat outcomes and probabilities linearly when making risky decisions (see for instance Camerer, 1995 or Starmer, 2000 for a survey). Insofar as our interactive decisions can be viewed as risky, it does not look implausible that individuals use similar decision rules, and deal with outcomes and probabilities in a non linear fashion, instead of maximizing expected-value as postulated by the basic model.

Formally, in CDR (2007), the SCs' probability of defection, given by $\alpha = \frac{(1-T)}{p}$, can be formulated as an explicit function of both the structure of the payoffs and the proportion p of SCs in the population (see Result (i), Section 2, *supra*). The fact that the probability α (empirically given by the frequency of defective choices) be much lower than theoretically expected can be imputed to either some overweighting of the probability p of meeting a SC, or undervaluation of $(1-T)$, or even both.

For instance, reading Graph 1 horizontally shows that the 30% level of cooperation that should be observed for p = 0.7 is actually observed for p = 0.45. This suggests that subjects behave *as if* they considered p as much higher than it actually is (they cooperate much more than they theoretically should), i.e. as if they *overweighted* it.

Similarly, assuming that the subjects deal with outcomes through some utility function u might help understand why they behave as they do when H varies. The gap between empirical and predicted behaviour is especially wide (reaching 50%), and the difference between the two increases then decreases as H grows. Since the impact of H on behaviour is examined for a given proportion of SCs, which is kept constant throughout the questions (p = 0.75), subjective probability weighting alone cannot account for the data. A more plausible explanation is that the subjects do not value their defection gain as much as they should, had they been expected-value maximizers. For instance, reading Graph 3 horizontally shows that the level of defection observed when H = 150 (resp. H = 190) corresponds to what was expected for H = 120 (resp. H = 130). This suggests that the subjects behave *as if* they systematically and significantly (by 60 in the most extreme case) undervalued either the unilateral defection gain H (i.e. $u(H) < H$) or the difference between H and the other elements of the matrix (being held constant). Such undervaluation of the defection gain may help understand why the subjects keep on cooperating while they no longer should.

Similarly, since the impact of L on behaviour is examined for a given proportion of SCs (namely p = 0.75), most of the large and varying gap between predicted and observed behaviour cannot be interpreted in terms of probability weighting alone. Reading Graph 4 horizontally shows that, for instance, the level of cooperation obtained for L = 50 € is identical to the theoretically predicted one for L = 85 €. This suggests that the subjects behave as if they systematically and significantly (by 35 in the most extreme case) overvalued either the unilateral cooperation gain L (i.e. $u(L) > L$) or the difference between L and the other elements of the matrix (being held constant). Such overvaluation of the cooperation gain may help understand why the subjects begin to cooperate while they should keep on defecting.

Conclusion

The experimental study described in this paper aimed at investigating the descriptive accuracy of a recent game theoretical model involving a Game of Chicken with agents' heterogeneity (CDR, 2007). The core and rather counterintuitive theoretical prediction we wanted to confront with real behaviour was the idea that self-interested people's proneness to cooperation tends to grow as their weight in the population increases. From a qualitative point of view, our data tend to support this prediction. Besides, the level of cooperation appears to be a decreasing (resp. increasing) function of the unilateral defection (resp. cooperation) gain.

Nevertheless, our data show a high level of cooperation. Although usual psychological, sociological and cognitive explanations should obviously not be excluded, our data suggest that attitude towards risk, and more specifically the way the subjects subjectively deal with outcomes and probabilities, may have also contributed to this strikingly high level of cooperation. Game theory assumes that the subjects are expected-value maximizers. Contemplating the possibility that individuals maximize their expected utility, or even some more sophisticated functional (as in Cumulative Prospect Theory, CPT, Tversky and Kahneman 1992 for instance), may help account for the gap between standard theoretical predictions and observed behaviour.

Unfortunately, our data do not allow us to ensure that the subjects actually dealt with outcomes and probabilities the way we suggest. In this respect, a future track for research would be to build an extension of CDR (2007), in which agents would be endowed with CPT preferences, thus allowed to deal with both outcomes and probabilities non linearly. In that extended framework, some new predictions could be drawn and confronted with observed behaviour in a more systematic manner.

NOTES

1. In that study, personality was captured through such indicators as sensations seeking or locus of control.
2. The pilot experiment was run using the two-stage version of the game, involving a first choice between entry and non entry, and, in case of entry, a second choice between cooperation and non cooperation. The resulting choice situations appeared to be too complex for the subjects, resulting in poor concentration and unreliable answers.
3. This probability corresponds to the probability $1-p$ to meet a UC (who always cooperates) added to the probability $(1-\alpha)\,p$ to meet a SC who chooses to cooperate.
4. It is a well-known result in standard game theory (when all players are assumed to be 'self-interested') that any Game of Chicken offers three equilibria, two of them involving pure strategies. In this case, one of the players initiates a conflict and the other one chooses to cooperate, leading the latter to be declared the « chicken » of the game. The last equilibrium involves completely mixed strategies, with each player assigning a positive probability to each of her strategies, so as to maximize her payoffs. So, $(1-\alpha) = L/L+H$ corresponds to the mixed-strategy equilibrium of the game when there is no heterogeneity in the population, i.e. when all the players are assumed to be 'self-interested'.
5. Indeed, even if only the 'deep structure' of the game is theoretically relevant for decision-making, there is a huge body of evidence confirming that the 'surface structure' (framing, wording, etc.) sharply influences the way people behave (Poppe, 2005; Wagenaar et al., 1988). In this respect, a socially-oriented framing might prevent a proper investigation of CDR (2007)'s predictions and confuse the data.
6. The first two questions actually provided guidance for the subjects. They were introduced after the pilot experiment showed that it was easier for the subjects to think of extreme values (which often induce simple decisions, due to dominance effects for instance) first.
7. Tables offer the advantage of simplicity and compacity, but they may also encourage the use of undesirable heuristics.
8. Actually, since the sessions were completely independent from each other, each of them had its own performance-based payment procedure.
9. Even though the ex ante probability of gain is quite low, the subjects appeared to be very sensitive to the payment procedure: they focused on the best outcome rather than on the probability of winning.

10. Of course, non-SC subjects were not excluded from the performance-based payment procedure. Neither were the subjects who produced erratic answers.
11. Global behaviour is actually almost entirely driven by the UCs' deterministic behaviour.

REFERENCES

Berg, J. J. W. Dickaut and K. A. McCabe 1995. 'Trust, Reciprocity and Social History.' *Games and Economic Behavior* 10: 122-142.

Bohnet, I. and R. Zeckhauser 2004. 'Trust, Risk and Betrayal.' *Journal of Economic Behavior and Organization* 55: 467-484.

Bolton, G. and A. Ockenfels 2000. 'A Theory of Equity, Reciprocity and Competition.' *American Economic Review* 90: 166-193.

Boone, C., B. de Brabander and A. van Witteloostuijn 1999. 'The Impact of Personality on Behavior in Five Prisoner's Dilemma Games.' *Journal of Economic Psychology* 20: 343-377.

Boone, C. and A. van Witteloostuijn, 1999. 'Competitive and Opportunistic Behavior in a Prisoner's Dilemma Game: Experimental Evidence on the Impact of Culture and Education.' *Scandinavian Journal of Management* 15: 333-350.

Brennan, G., L. G. Gonzalez, W. Güth and M. V. Levati 2007. 'Attitudes toward Private and Collective Risk in Individual and Strategic Choice Situations.' *Journal of Economic Behavior and Organization*. in press.

Buskens, V. 2003. 'Trust in triads: effect of exit, control and learning.' *Games and Economic Behavior*. 42: 235-252.

Cabon-Dhersin, M-L. and S. V. Ramani 2007. 'Opportunism, Trust and Cooperation: A Game Theoretic Approach with Heterogeneous Agents.' *Rationality and Society* 79: 203-228.

Camerer, C. 2003. *Behavioral Game Theory: Experimental Studies of Strategic Interaction*. Princeton: Princeton University Press.

Camerer, C. 1995. 'Individual Decision Making.' in *Handbook of Experimental Economics*. J. Kagel and A. E. Roth. (eds). Princeton: Princeton U. Press.

Carraro, C. and D D. Siniscalco 1993. 'Strategies for the International Protection of the environment.' *Journal of public Economics* 52: 309-328

Charness, G. and M. Rabin 2002. ‘Understanding Social Preferences with simple tests.’ *Quartely Journal of Economics*. 117: 817-869.

Dawes, R. M., A. J. Van de Kragt and J. M. Orbell. 1988.’Not Me or Thee but We: The importance of Group Identity in Eliciting Cooperation in Dilemma Situations: Experimental Manipulations.’ *Acta Psychologica* 68: 83-97.

Dolbear, F.T. and L.B. Lave 1966. ‘Risk Orientation as a Predictor in the Prisoner’s Dilemma.’ *The Journal of Conflict Resolution* 10: 506-515.

Eckel, C.C. and R. K. Wilson 2004. ‘Is Trust a Risky Decision?.’ *Journal of Economic Behavior and Organization* 55: 447-465.

Falk, A. and U. Fischbacher 2006. ‘A Theory of Reciprocity.’, *Games and Economic Behavior* 54: 293-315.

Fehr, E. and U. Fischbacher 2002. ‘Why Social Preferences Matter-the impact of non-selfish motives on competition, cooperation and incentives.’ *The Economic Journal* 12:xx.

Fehr, E. and S. Gächter S. 2000. ‘Fairness and Retaliation: The Economics of Reciprocity.’ *Journal of Economic Perspectives* 14: 159-181.

Fehr, E. and K. Schmidt 1999. ‘A theory of Fairness, Competition, and Cooperation.’ *Quaterly Journal of Economics*. 3: 817-868.

Frank, R. H. 1987. ‘If homo economicus could choose his own utility function, would he choose one with a conscience?.’ *American Economic Review*. 77: 693-704.

Friedman, M., & Savage L. J. 1948. ‘The Utility Analysis of Choices Involving Risk.’ *Journal of Political Economy* 56: 279-304.

Gächter, S., B. Herrmann, and C. Thöni 2004. ‘Trust, Voluntary Cooperation, and Socio-Economic Background: Survey and Experimental Evidence.’ *Journal of Economic Behavior and Organization*. 55: 505-531.

Goeree, J.K. and C. A. Holt 2001. ‘Ten Little Treasures of Game Theory and Ten Intuitive Contradictions.’ *American Economic Review* 91: 1402-1422.

Güth, W, P. Ockenfels, and M. Wendel 1997. ‘Cooperation Based on Trust: An Experimental Investigation.’ *Journal of Economic Psychology* 18: 18-43.

Hardin, R. 1971. ‘Collective Action as an Agreeable N-Prisoners' Dilemma.’ *Science* 16: 472-481.

Kollock, P. 1998a. ‘Social Dilemmas: The Anatomy of Cooperation.’ *Annual Review of Sociology* 24:183-214.

Kollock, P. 1998b. ‘Transforming Social Dilemmas: Group Identity and Cooperation.’ In *Modelling Rational and Moral Agents,* ed. P. Danielson, pp. 186-210.

Levine, D. 1998. ‘Modeling Altruism and Spitefulness in Experiments.’ *Review of Economic Dynamics* 1: 593-622.

McClintock, C.G. and W. B. G. Liebrand. 1988. Role of Interdependance Structure, Individual Value Orientation, and Another’s Strategy in Social Decision Making: A transformational Analysis.’ *Journal of Personality and Social Psychology* 55: 396-409.

Neugebauer T., A. Poulsen and A. Schram A. 2008. ‘Fairness and Reciprocity in the Hawk-Dove Game.’ *Journal of Economic Behavior and Organization*. 66(2): 243-250.

Poppe, M. 2005. ‘The Specificity of Social Dilemma Situations.’ *Journal of Economic Psychology* 26: 431-441.

Rabin, M. 1993. ‘Incorporating Fairness into Game Theory and Economics.’ *American Economic Review* 83: 1281-1302.

Ridley, M. 1997. ‘The Origins of Virtue: Human instincts and the Evolution of Cooperation’. New York: Viking Penguin.

Sabater-Grande, G. and G. Georgantzis 2002. ‘Accounting for Risk Aversion in Repeated Prisoners’ Dilemma Games: An Experimental Test.’ *Journal of Economic Behavior and Organizatio.* 48: 37-50.

Sherman, R. 1969. ‘Risk Attitude and Cost Variability in a Capacity Choice Experiment.’ *Review of Economic Studies* 36: 453-466.

Simpson, B. 2006. ‘Social Identity and Cooperation in Social Dilemmas.’ *Rationality and Society* 18: 443-470.

Snyder, G. H. 1971, ‘"Prisoner's Dilemma" and "Chicken" Models in International Politics.’ *International Studies Quarterly* 15: 66-103.

Starmer, C. 1999. ‘Experiments in Economics: Should We Trust the Dismal Scientists in White Coats?’ *Journal of Economic Methodology*, 6(1): 1-30.

Starmer, C. 2000. ‘Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice Under Risk.’ *Journal of Economic Literature* 38: 332-382.

Stone, R. W. 2001. ‘The Use and Abuse of Game Theory in International Relations: The Theory of Moves.’ *The Journal of Conflict Resolution* 45: 216-244.

Thogersen, J. 2008. ‘Social Norms and Cooperation in Real-Life Social Dilemmas.’ *Journal of Economic Psychology*, in press.

Turner, J. C., M. A. Hogg, P.J. Oakes, S. D. Reicher, and M. S. Weetherell. 1987. *Rediscovering the Social Group: A self-Categorization Theory.* Oxford: Basil Blackwell.

Tversky, A., and Kahneman D. 1992. ‘Advances in Prospect Theory: Cumulative Representation of Uncertainty’. *Journal of Risk and Uncertainty* 5: 297-323.

Yaari, M. 2001. ‘A Credit Market à la David Hume.’ Economic Working Papers. School of Social Science. Institute for Advanced Study, Princeton. New Jersey.

Wagenaar, W. A., G. Keren and S. Lichtenstein 1988. ‘Islanders and Hostages: Deep and Surface Structures of Decision Problems.’ *Acta Psychologica* 67: 175-188.

Ward, H. 1993. ‘Game Theory and the Politics of the Global Commons,’ *The Journal of Conflict Resolution* 37: 203-235.

Wit A. P. and H. A. M. Wilke. 1992. ‘The Effect of Social Categorization on Cooperation in Three Types of Social Dilemmas’. *Journal of Economic Psychology* 13: 135-151.

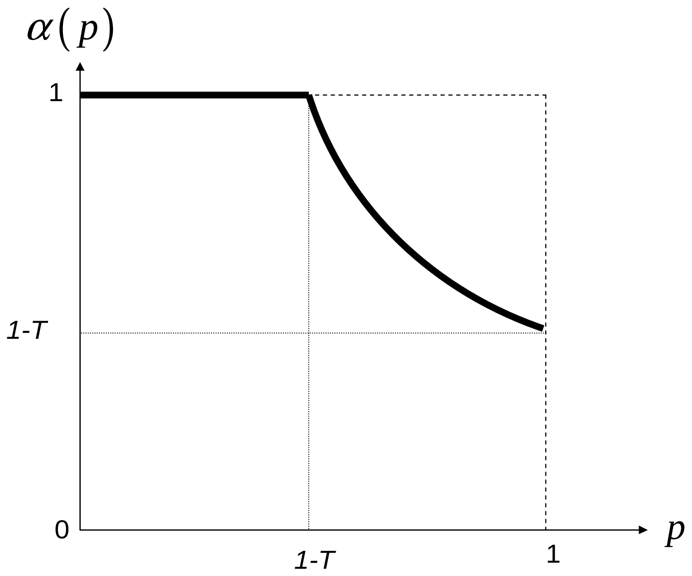
Yamagishi, T. and T. Kiyonari. 2000. ‘The group as the Container of Generalized Reciprocity.’ *Social Psychology Quartely* 63: 116-32.

FIGURES AND GRAPHICS.

Figure 1. A typical matrix of the Game of Chicken ($H > X > L > Y$)

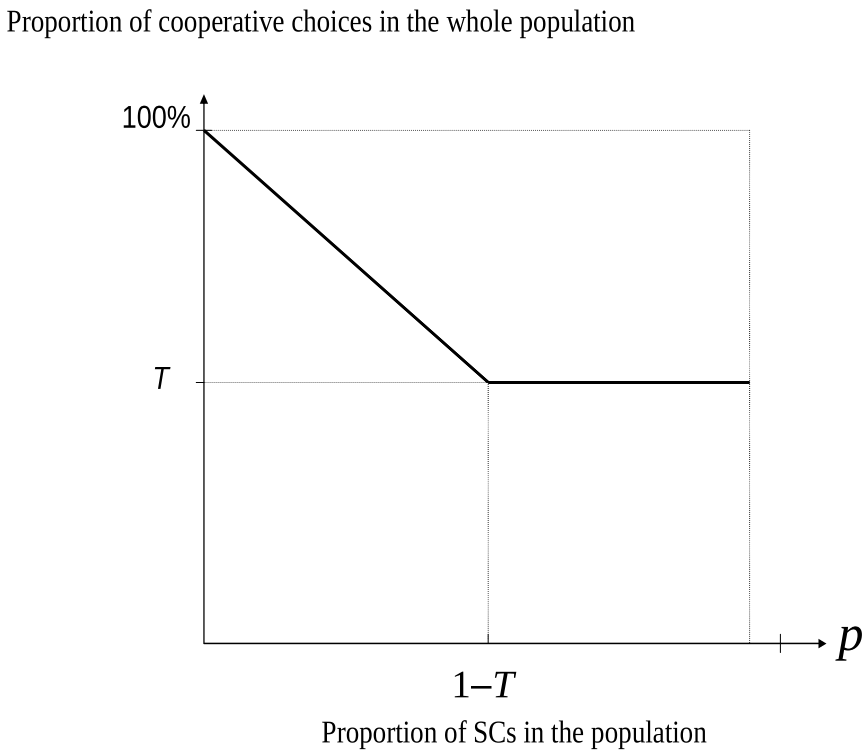
		Player *j*	
		$1-\alpha$ **c**	α **d**
Player *i*	$1-\alpha$ **c**	$X ; X$	$L ; H$
	α **d**	$H ; L$	$Y ; Y$

Figure 2. The probability of defection depending on the proportion of SCs in the population[1]



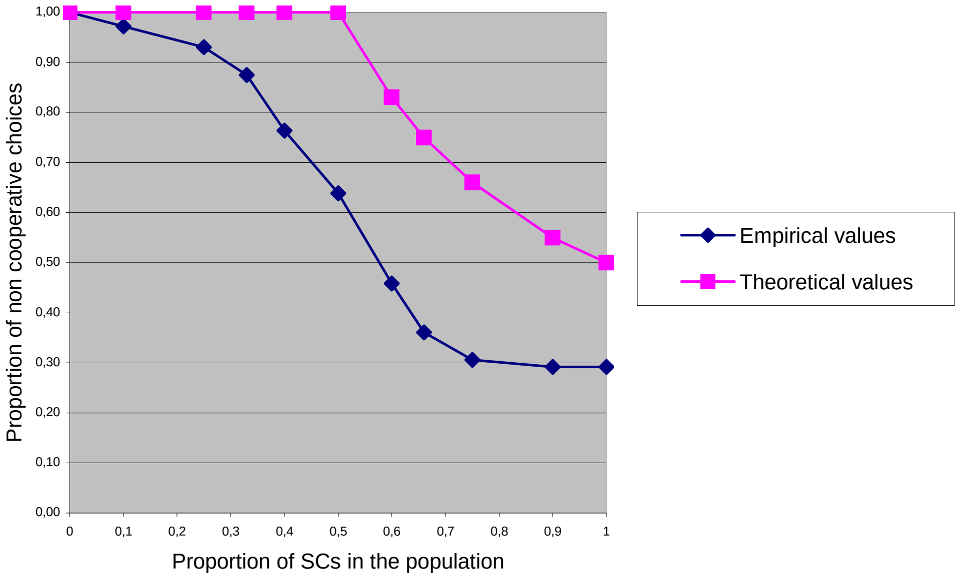
[1] We consider here the following payoff structure: H = 120, L = 70, X = 100 and Y = 50.

Figure 3. The proportion of cooperative choices in the whole population depending on the proportion of SCs in the population[2].

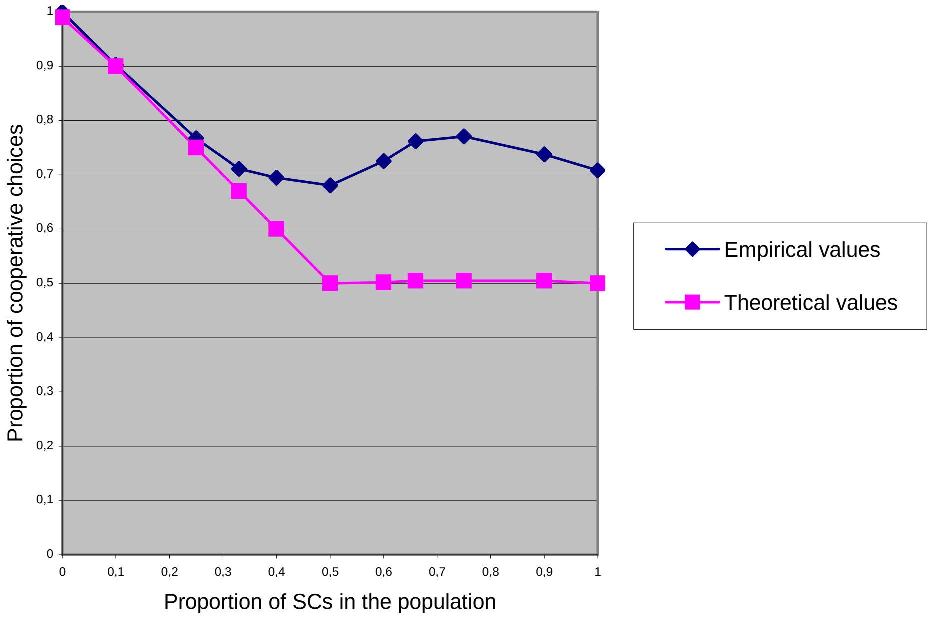


[2] We consider here the following payoff structure: H = 120, L = 70, X = 100 and Y = 50.

Graph 1. Proportion of non cooperative choices among the SCs depending on the proportion of SCs in the population.

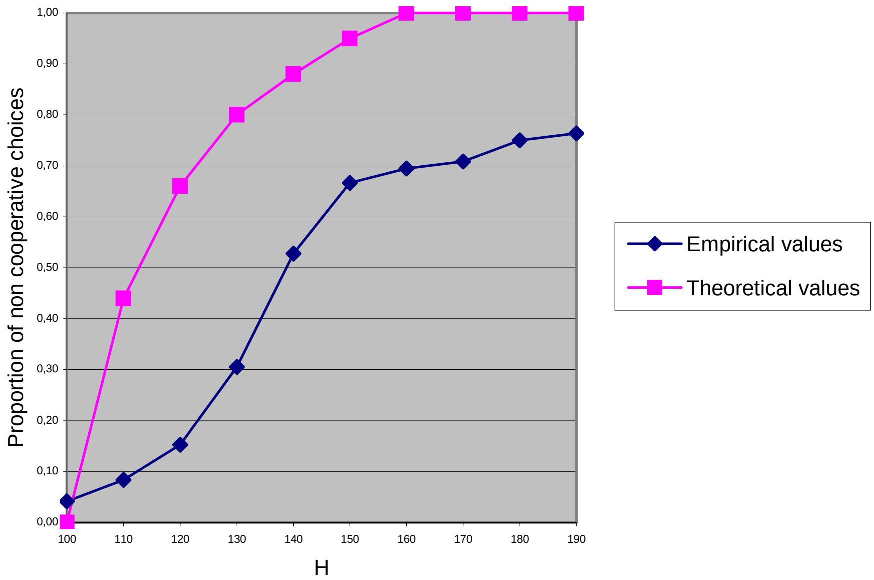


Graph 2. Proportion of cooperative choices in the whole population depending on the proportion of SCs in the population.



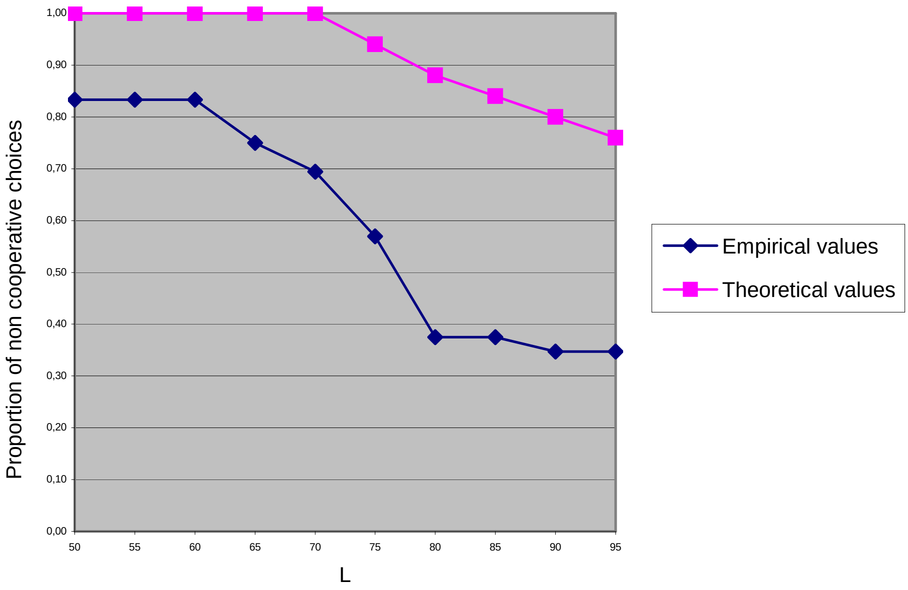
Graph 3. Proportion of non cooperative choices among the SCs

depending on unilateral defection gain H.



Graph 4. Proportion of non cooperative choices among the SCs

depending on unilateral cooperation gain L



Appendix A: A typical choice situation (drawn from the first part of the questionnaire)

You are facing a partner. You both have two available options: play **Red** or play **Blue**. You do not know what your partner is going to do, but your gain depends on both your own choice and hers.
The matrix below gives the different choices for you and your partner, as well as the corresponding gains. Your gains are in **bold**:

		The choice of my partner	
		Red	Blue
My choice	**Red**	**100 €**; 100 €	**70 €**; 120 €
	Blue	**120 €**; 70 €	**50 €**; 50 €

Question 1.1 :
You have **100%** chances to meet a partner of **type (b)** who plays either Red or Blue, depending on her potential gains as well as on what she thinks you are going to play.
Which colour do you choose?

□ **Red**
□ **Blue**

Question 1.2 :
You have **100%** chances to meet a partner of **type (a)** who always plays Red.
Which colour do you choose?

□ **Red**
□ **Blue**

Question 1.3 :
We are now in the general case. Your partner has:

- **X%** chances to be of **type (a)**, in which case she always plays Red,
  and
- **(100-X)%** chances to be of **type (b)**, in which case she plays either Red or Blue, depending on her potential gains as well as on what she thinks you are going to play.

Which colour do you choose for the different values of X that are given in the table below? Just tick the appropriate box (**Red** or **Blue**) for each of these values.

X% chances to meet a partner of type (a) (who always plays Red) → ***Your choice*** ↓	**0%**	**10%**	**25% (or 1/4)**	**33% (or 1/3)**	**40%**	**50%**	**60%**	**66% (or 2/3)**	**75% (or 3/4)**	**90%**	**100%**
Red											
Blue											

Question 1.1. ↑ (0% column)

Question 1.2. ↑ (100% column)